

Full valency and the position of enclitics in the Old Czech

Radek Čech

University of Ostrava, Faculty of Arts
Department of Czech Language
Czech Republic
cechradek@gmail.com

Pavel Kosek

Masaryk University, Faculty of Arts
Department of Czech Language
Czech Republic
kosek@phil.muni.cz

Olga Navrátilová

Masaryk University, Faculty of Arts
Department of Czech Language
Czech Republic
olga@phil.muni.cz

Ján Mačutek

Masaryk University, Faculty of Arts
Department of Czech Language
Czech Republic,
and
Comenius University in Bratislava
Faculty of Mathematic, Physics and Infor-
matics
Department of Applied Mathematics and
Statistics
Slovakia
jmacutek@yahoo.com

Abstract

The paper is focused on the analysis of the relationship between the full valency of the predicate and the position of enclitics in the clause. For this analysis, ones of the oldest Old Czech prose texts were used. We set up the hypothesis - the higher the full valency of the predicate, the lower the probability of the occurrence of the enclitic after the initial phrase of the clause – and test it. The hypothesis was corroborated only for narrative texts. In the case of poetic texts, the hypothesis was rejected.

1 Introduction

Enclitics are language units with a variety of specific grammatical characteristics that have attracted linguists for decades. Despite the fact that a huge number of methods and theoretical approaches were applied to study of this phenomenon, some fundamental questions are still open. Among others, an empirical diachronic description and an explanation of the historical development of enclitics that is based on a larger amount of language material and interpreted from the point of view of the quantitative linguistics remain rather an unexplored field. The reasons are obvious: language material is accessible only with difficulties (in the majority of cases it must be transcribed from a manuscript); an annotation must be performed manually; for the oldest periods, a limited number of texts is available, etc. However, an analysis of the historical development of any language property (or unit) often brings knowledge that substantially enhances an understanding of the phenomena under study. Therefore, in a recent series of papers several properties of enclitics in Old Czech and their historical development were explored (cf. Kosek et al., 2018a, 2018b, 2018c, 2018d), with the aim of obtaining a diachronic perspective of their characteristics. This paper represents a further step in this endeavour. Specifically, we analyse the relationship between the position of the pronominal enclitic in the clause and the so-called full valency (Čech et al., 2010) of the clause predicate. We assume that the full valency (for details, see Section 4) is one of factors which significantly influence the position of the enclitic. Therefore, we set up the following hypothesis:

The higher the full valency of the predicate, the lower the probability of the occurrence of the enclitic after the initial phrase of the clause.

The position of the enclitic immediately after the initial phrase (hereafter, this position will be called the postinitial position, abbreviated as 2P) is considered, according to the Wackernagel's Law (Wackernagel, 1892), the basic position of this unit in the clause, cf. the position of reflexive "sě" in the sentence (1).

(1)

[Co] *sě tobě vidí, Šimone?*

what_{NOM} REFL_{ACC} see_{3.PS.SG.PRAES}

'What is thy opinion, Simon?'

Bible olomoucká (BiblOl) Mathew 17,24

In previous studies (Kosek et al., 2018c; Čech et al., 2019; Kosek et al., 2019) it was shown that there are several factors which move the enclitic to a position which is different from 2P: for instance, the length of the initial phrase, the style, and the impact of the original Latin pretext. All these factors decrease the probability of the occurrence of the enclitic in the 2P position. This study is based on the assumption that the full valency is another factor which should lead to a similar result. Reasons for this assumption are summarized in Section 4 where the notion of full valency is introduced in detail. Statistical methods applied in this paper require certain amount of data to be reliable, therefore, only the most frequent pronominal enclitic in Old Czech, i.e. the enclitic "sě" (accusative reflexive) is analysed. Data from the Olomouc Bible (Bible olomoucká, BiblOl) and Litoměřice-Třeboň Bible (Bible litoměřicko-třeboňská, BiblLitTřeb), which represent the oldest complete Czech Bible translation, are used.

2 Word order of enclitics in Old Czech

For the purpose of this study, we determine two positions of the enclitic (E) in Old Czech:

a) the postinitial position, schematically

[I][E][]*

where symbol [I] represents the initial phrase of the clause and symbol []* represents any consequent syntactic unit(s) of the clause (including the empty unit, i.e. the clause can end with the enclitic). The initial phrase can be represented by one or more words, cf. sentence (1) and (2) respectively.

(2)

[*toho věku*] *sě jemu porodil Isák*

that_{GEN.F.SG} age_{GEN.F.SG} REFL_{ACC} him_{DAT.M.SG} born_{PART.PRET.ACT.NOM.SG.M} Isaac_{NOM.M.SG}

'And as Abraham was a hundred years old, his son Isaac was born to him.'

BiblOl Genesis 21,5

b) non-postinitial positions, schematically

[I][[]*][E][[]*

cf. sentence (3)

(3)

[*Volanie Sodomských a Gomorrejských*] *rozмноžilo sě jest*

outcry_{NOM.N.SG} sodom_{ADJ.GEN.M.PL} and gomorrha_{ADJ.GEN.M.PL} multiply_{PART.PRET.ACT.N.SG} REFL_{ACC}
be_{AUX.PRET.3.SG}

'The cry of Sodom and Gomorrha is multiplied'

BiblOl Genesis 18,20

3 Language material

For the analysis, some books from the Olomouc Bible (Bible olomoucká, BiblOl) and one book (Acts) from Litoměřice-Třeboň Bible (Bible litoměřicko-třeboňská, BiblLitTřeb) were used. These Bibles originate from the beginning of 15th century (however, it is considered to be copied from missing older translation from 1360, cf. Kyas, 1997; Vintr, 2008) and can be ranked among the oldest Old Czech prose texts (older texts, from the first half of the 14th century, are poetic, and they cannot be used to observe word order characteristics). Since our long-term aim is an analysis of the historical development of the word order characteristics of enclitics, the use of one of the oldest texts seems to be a proper choice - the result of this study can be, afterwards, compared with the results based on later Czech Bible translations.

All the phenomena under the study must be annotated manually, therefore, only eight books from the Bible were analysed. Specifically, four books from the Old Testament and four books from the New Testament were chosen: Genesis (Gen), Isaiah (Is), Job (Job), Ecclesiastes (Ecc), Gospel of St. Matthew (Mt), Gospel of St. Luke (Lk), Acts (Act), and Revelation (Rev).

4 Full valency and word order of enclitics

The notion of full valency (FV) was introduced to linguistics by Čech et al. (2010) and was elaborated by Vincze (2014) and Čech et al. (2015). The FV approach is a reaction to the absence of reliable criteria for distinguishing obligatory arguments (complements) and non-obligatory arguments (optional adjuncts). FV does not distinguish between obligatory arguments and non-obligatory ones. Thus, all directly dependent units of the predicate which occur in the actual language usage comprise its full valency frame.

A higher FV of the predicate means a higher complexity¹ of the clause (at least at this level of the syntactic tree, i.e. at the root of the clause and its direct dependents). We assume that the higher complexity is the factor which increases the probability that the Wackernagel's Law is "violated" for the following reason. The occurrence of the enclitic in the 2P position often means that the enclitic is not in the position adjacent to its syntactically superior word. Further, a more complex the clause structure increases the difficulty of processing the clause structure cognitively, especially when it contains distant dependency relations. Consequently, the tendency to put the enclitic next to its syntactically superior word instead to the 2P position should be positively correlated with the complexity of the clause. Of course, the complexity of the clause could be, in the ideal case, determined from the property of the entire clause structure. However, the character of the language material, which must be annotated manually, forced us to focus exclusively on the FV as the measure of the clause complexity (i.e., only the highest levels of syntactic trees are taken into account). Admittedly, this approach has its limitations and more comprehensive characteristics of the syntactic complexity will have to be applied in future research, but results achieved indicate that the positions of the enclitics are indeed influenced also by syntactic properties of the clause of which they are part.

5 Results

The relationship between the FV and the proportion of enclitics in the 2P position is presented in Table 1 and Figure 1. Here, all data are merged together, i.e. these results represent property of the corpus comprising eight Biblical books (see Section 2). Since some FV sizes do not contain enough instances for a proper evaluation of the data (e.g., there are only four clauses with enclitics for which the FV attains the value of seven), we decided to pool the adjacent bins so that each bin contains at least ten instances. In Tables 2 and 3, the FV size expresses the weighted arithmetic mean, with frequencies being the weights.

¹ In quantitative linguistics it is usual to measure complexity of a syntactic structure as the number of its constituents (e.g. Köhler, 2012, p. 145). For other approaches to syntactic complexity see e.g. Miestamo et al. (2008) or Givón and Shibatami (2009).

FV	2P	non-2P	proportion of 2P
2	2	18	0.1
3	133	75	0.64
4	81	117	0.41
5	47	49	0.49
6.13	14	18	0.44

Table 1: The size of the full valency (FV), number of enclitics in the postinitial (2P) and non-postinitial position (non-2P), and proportion of the 2P in all chosen biblical books.

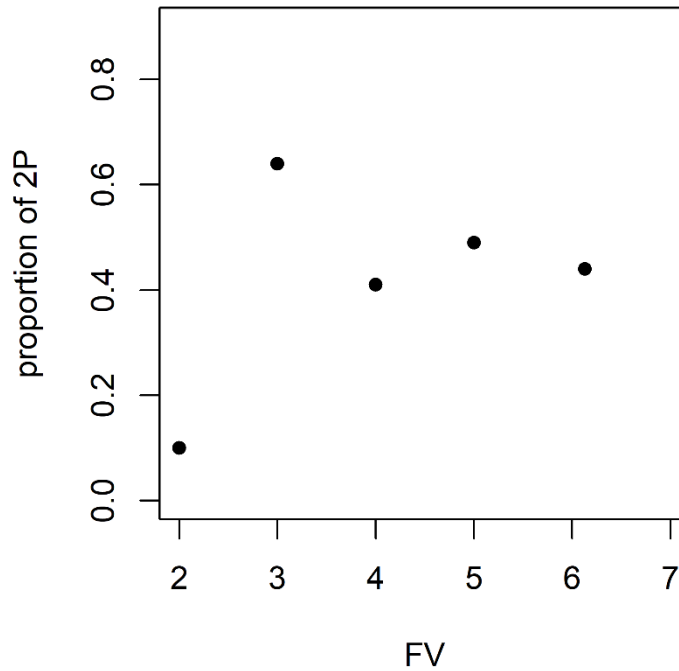


Figure 1: The relationship between the full valency (FV) and the proportion of enclitics in the postinitial (2P) position in all chosen biblical books

It is obvious (see Table 1 and Figure 1) that there is no tendency corresponding to our prediction from Section 1. On the contrary, clauses with the lowest FV have the lowest proportion of the 2P positions of enclitics which is a direct contradiction to the hypothesis. However, it is known that the distribution of particular positions of enclitics is significantly influenced by the style (Kosek et al., 2018c). As Biblical books fundamentally differ with respect to style, we studied also the relationship between the FV and the position of the enclitic separately in individual books. The results are presented in Table 2 and Figure 2.

The analysis of individual books brings rather a different picture. We can see that results from Act, Mt, Lk, and Gen corroborate the hypothesis, while results from Job and Ecc falsify it. As for Is and Rev, there are not enough data for a conclusion. At the first sight, it seems that there are differences between narrative texts (i.e., Act, Mt, Lk, and Gen) and poetic texts (i.e., Job and Ecc). In the case of poetic texts, their specific character can be a reason why the hypothesis is rejected – the author must fulfil some conditions to fit the rules of poetry, which can influence (or violate) the mechanism underlying the hypothesis.

VF_{Act}	$2P_{Act}$	non- $2P_{Act}$	proportion of $2P_{Act}$	VF_{Lk}	$2P_{Lk}$	non- $2P_{Lk}$	proportion of $2P_{Lk}$
2	2	18	0.1	2.86	20	16	0.56
3	133	75	0.64	4	18	18	0.5
4	81	117	0.41	5.16	8	11	0.42
5	47	49	0.49				
6.13	14	18	0.44				

VF_{Mt}	$2P_{Mt}$	non- $2P_{Mt}$	proportion of $2P_{Mt}$	VF_{Gen}	$2P_{Gen}$	non- $2P_{Gen}$	proportion of $2P_{Gen}$
2.95	15	6	0.71	2.91	20	12	0.63
4	7	12	0.37	4	13	19	0.41
5.64	5	9	0.36	5.25	11	13	0.46

VF_{Job}	$2P_{Job}$	non- $2P_{Job}$	proportion of $2P_{Job}$	VF_{Ecc}	$2P_{Ecc}$	non- $2P_{Ecc}$	proportion of $2P_{Ecc}$
2.92	23	15	0.61	2.93	34	27	0.56
4	11	21	0.34	4	12	27	0.31
5.25	11	9	0.55	5	9	2	0.82

Table 2: The size of the full valency (FV), number of enclitics in the postinitial (2P) and non-postinitial position (non-2P), and proportion of the 2P in individual Biblical books.

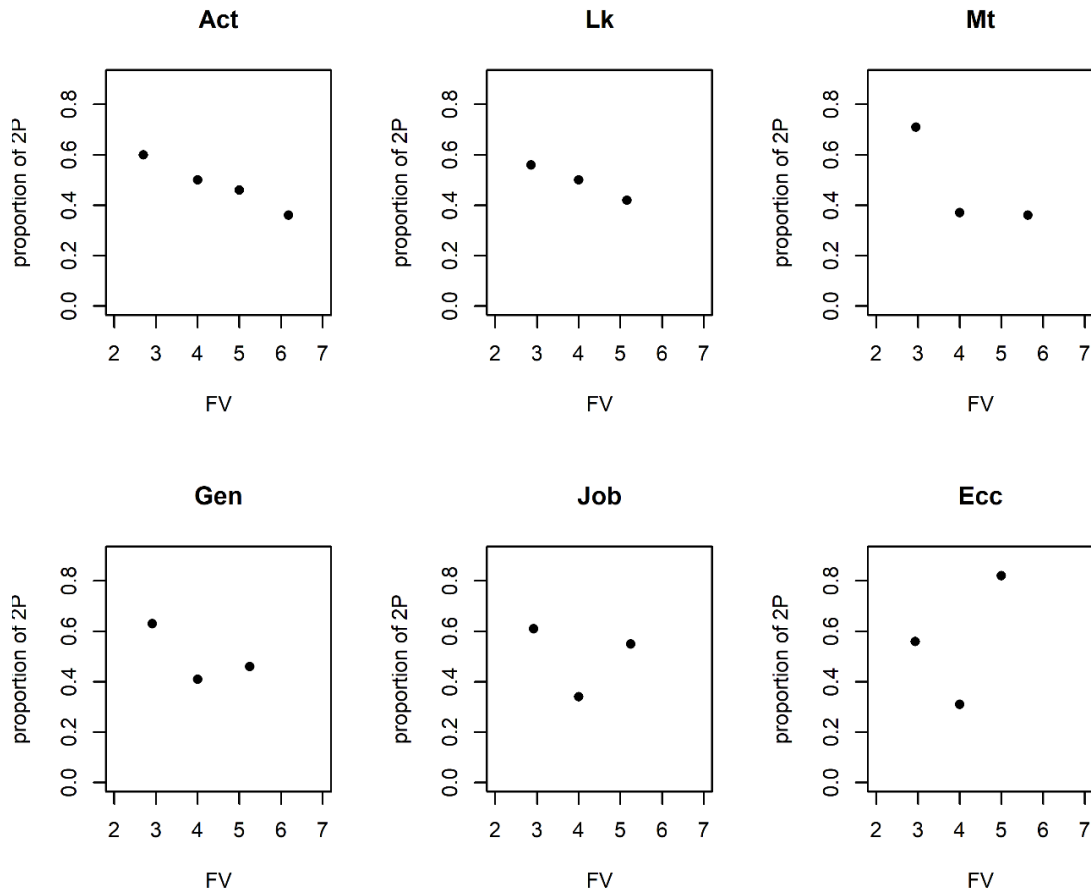


Figure 2: The relationship between the full valency (FV) and the proportion of enclitics in the postinitial (2P) position in individual Biblical books.

6 Conclusions

The study brings some important findings. First, even though the hypothesis was falsified when it was tested on both poetic texts and a corpus consisting of eight Biblical books, we do not reject the hypothesis generally. We assume that the poetic character of texts can be interpreted as a border condition which restricts the validity of the hypothesis. Further, it was revealed that mixing texts is another factor that can influence the outcome of hypothesis testing significantly. A mixture of different texts (e.g. with respect to their genre or style) means that particular mechanisms can "fight" each other and, as a consequence, their influence can be weakened (or it can even disappear). Finally, it must be emphasized that this paper is the first attempt to test this hypothesis. Needless to say, further research is necessary in this research field.

Acknowledgements

This study was supported by the project *Development of the Czech pronominal (en)clitics* (GAČR GA17-02545S).

References

- Radek Čech, Petr Pajas, and Ján Mačutek. 2010. Full Valency. Verb Valency without Distinguishing Complements and Adjuncts. *Journal of Quantitative Linguistics*, 17:291-302.
- Radek Čech, Ján Mačutek, and Michaela Koščová. 2015. On the relation between verb full valency and synonymy. In Eva Hajičová and Joachim Nivre (eds.), *Proceedings of the Third International Conference on Dependency Linguistics (Depling 2015)*:68–73. Uppsala University, Uppsala.
- Radek Čech, Pavel Kosek, Olga Navrátilová, and Ján Mačutek. 2019. On the impact of the initial phrase length on the position of enclitics in the Old Czech. *Proceedings of QUALICO 2018* (submitted).
- Talmy Givón and Masayoshi Shibatani. 2009. *Syntactic Complexity: Diachrony, Acquisition, Neuro-Cognition, Evolution*. Benjamins, Amsterdam/Philadelphia.
- Pavel Kosek, Radek Čech, Olga Navrátilová, and Ján Mačutek. 2018a. On the Development of Old Czech (En)clitics. *Glottometrics*, 40:51-62.
- Pavel Kosek, Olga Navrátilová, Radek Čech, and Ján Mačutek. 2018b. Word Order of Reflexive 'sě' in Finite Verb Phrases in the First Edition of the Old Czech Bible Translation. (Part 1). *Studia Linguistica Universitatis Iagellonicae Cracoviensis*, 135(3):177-188.
- Pavel Kosek, Olga Navrátilová, Radek Čech, and Ján Mačutek. 2018c. Word Order of Reflexive 'sě' in Finite Verb Phrases in the First Edition of the Old Czech Bible Translation. (Part 2). *Studia Linguistica Universitatis Iagellonicae Cracoviensis*, 135(3):189-200.
- Pavel Kosek, Olga Navrátilová, and Radek Čech. 2018d. Slovosled staročeských pronominálních enklitik závislých na VF ve staročeské bibli 1. redakce. *SLAVIA časopis pro slovanskou filologii*, 87(1-3):189-204.
- Pavel Kosek, Radek Čech, and Olga Navrátilová. 2019. The influence of the Latin pretext on the word order of pronominal enclitics in the Bible of Olomouc and the Bible of Litoměřice-Třeboň. *Proceeding of the 3rd Diachronic Slavonic Syntax conference DSSL* (submitted).
- Reinhard Köhler. 2012. *Quantitative Syntax Analysis*. De Gruyter, Berlin/Boston.
- Vladimír Kyas. 1997. *Česká Bible v dějinách národního písemnictví*. Praha: Vyšehrad.
- Matti Miestamo, Kaius Sinnemäki, and Fred Karlsson (eds.). 2008. *Language Complexity: Typology, Contact, Change*. Benjamins, Amsterdam/Philadelphia.
- Veronika Vincze. 2014. Valency frames in a Hungarian corpus. *Journal of Quantitative Linguistics*, 21(2):153–176.
- Josef Vintr. 2008. Bible (staroslověnský překlad, české překlady). In Luboš Merhaut et al. (eds.), *Lexikon české literatury* [vol. 4/2: U–Ž; Dodatky A–Ř]: 1882–1887. Academia, Praha.
- Jacob Wackernagel. 1892. Über ein Gesetz der indogermanischen Wortstellung. *Indogermanische Forschungen*, 1(1): 333-436.